

## Cover Page for Statistical Analysis Plan

Sponsor name:	Novo Nordisk A/S
NCT number	NCT02648204
Sponsor trial ID:	NN9535-4216
Official title of study:	Efficacy and safety of semaglutide versus dulaglutide as add-on to metformin in subjects with type 2 diabetes
Document date:	30 November 2017

## 16.1.9 Documentation of statistical methods

### List of contents

Statistical analysis plan.....	<a href="#">Link</a>
--------------------------------	----------------------

## Statistical Analysis Plan

**Trial ID: NN9535-4216**


**SUSTAIN 7**

**Efficacy and safety of semaglutide versus dulaglutide as add-on to metformin in subjects with type 2 diabetes**

*Redacted statistical analysis plan  
Includes redaction of personal identifiable information only.*

**Trial phase: 3b**

**Author:**

, Trial Statistician  
Biostatistics

This confidential document is the property of Novo Nordisk. No unpublished information contained herein may be disclosed without prior written approval from Novo Nordisk. Access to this document must be restricted to relevant parties.

## Table of contents

	<b>Page</b>
<b>Table of contents .....</b>	<b>2</b>
<b>List of abbreviations .....</b>	<b>3</b>
<b>1 Introduction .....</b>	<b>5</b>
1.1 Trial information .....	5
1.2 Scope of the statistical analysis plan .....	5
<b>2 Statistical considerations .....</b>	<b>6</b>
2.1 General considerations .....	6
2.2 Primary estimand .....	6
2.3 Missing data considerations .....	7
2.4 Sample size calculation .....	7
2.5 Definition of analysis sets .....	9
2.6 Data selections and observations periods .....	10
2.7 Primary endpoint .....	11
2.8 Hypotheses tested for the primary endpoint .....	13
2.9 Multiplicity and criteria for confirming hypotheses .....	13
2.10 Sensitivity analyses .....	14
2.10.2 Sensitivity analyses addressing non-inferiority .....	15
2.10.3 Other sensitivity analyses .....	15
2.10.4 Assessment of sensitivity analyses .....	16
2.11 Secondary endpoints .....	17
2.11.1 Confirmatory secondary endpoint .....	17
2.11.2 Confirmatory secondary hypothesis .....	17
2.12 Supportive secondary endpoints .....	18
2.12.1 Efficacy endpoints .....	18
2.12.2 Safety endpoints .....	19
2.12.3 Adverse Events .....	21
2.13 Health economics and/or patient reported outcomes .....	24
<b>3 Changes to the statistical analyses planned in the protocol .....</b>	<b>25</b>
<b>4 References .....</b>	<b>26</b>

## List of abbreviations

AACE	American Association of Clinical Endocrinologists
ADA	American Diabetes Association
AE	Adverse event
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
BG	Blood glucose
BMI	Body mass index
BW	Body weight
CHF	Congestive Heart Failure
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
CI	Confidence interval
CV	Cardiovascular
CV	Coefficient of variation
CTR	Clinical trial report
DTSQ	Diabetes Treatment Satisfaction Questionnaire
EAC	Event adjudication committee
ECG	Electrocardiogram
eCRF	Electronic case report form
EoT	End of text
EOT	End of treatment
FAS	Full analysis set
FDA	US Food and Drug Administration
FPG	Fasting plasma glucose
GFR	Glomerular filtration rate
HbA <sub>1c</sub>	Glycosylated haemoglobin
HDL	High density lipoprotein
HOMA-B	Homeostasis model assessment
HR	Hazard ratio
ITT	Intention-to-treat
IV/WRS	Interactive voice/web response system
LDL	Low density lipoprotein
LLOQ	Lower limit of quantification
LOCF	Last observation carried forward
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
MMRM	Mixed model for repeated measures
PP	Per protocol
PRO	Patient reported outcome

PYO	Patient years of observation time
SAE	Serious adverse event
SAP	Statistical analysis plan
SAS	Safety analysis set
s.c.	subcutaneous
SD	Standard deviation
SE	Standard error
SEM	Standard error of mean
SMPG	Self-measured plasma glucose
SPS	Statistical Programming Specification
T2D	type 2 diabetes
TD	Treatment difference
ULOQ	Upper limit of quantification

# 1 Introduction

## 1.1 Trial information

This trial is 40-week randomised, open-label, active-controlled, parallel group, multicentre, multinational, four-armed trial to evaluate efficacy and safety of semaglutide versus dulaglutide as add-on to metformin in subjects with type 2 diabetes.

### Primary objective

To compare the effect of once-weekly dosing of two dose levels of subcutaneous semaglutide (0.5 mg and 1.0 mg) versus once-weekly dosing of two dose levels of subcutaneous dulaglutide (0.75 mg and 1.5 mg) on glycaemic control in subjects with type 2 diabetes on a background treatment with metformin.

### Key secondary objective

To compare the effect of once-weekly dosing of two dose levels of subcutaneous semaglutide (0.5 mg and 1.0 mg) versus once-weekly dosing of two dose levels of subcutaneous dulaglutide (0.75 mg and 1.5 mg) in subjects with type 2 diabetes on a background treatment with metformin with regards to body weight control, blood pressure, patient reported outcomes, and safety and tolerability.

### Trial design

Subjects with type 2 diabetes inadequately controlled with metformin alone will after approximately 2 weeks screening period be randomised in a 1:1:1:1 manner to receive either a dose of 0.5 mg or 1.0 mg of semaglutide once-weekly or 0.75 mg or 1.5 mg of dulaglutide once-weekly. After the treatment period of approximately 40 weeks in total, all subjects enter a follow-up period of 5 weeks ended by a follow-up phone contact. Total trial duration for the individual subjects will be approximately 47 weeks. A planned total number of 1196 subjects will be randomised. For further details, see protocol for trial NN9535-4216.

## 1.2 Scope of the statistical analysis plan

This statistical analysis plan (SAP) is based on the protocol for trial NN9535-4216 “Efficacy and safety of semaglutide versus dulaglutide as add-on to metformin in subjects with type 2 diabetes”, version 2.0 (4 November 2016), and includes more detailed procedures for executing the statistical analyses of the primary and secondary endpoints. Additional analyses have been added in this SAP as compared to the protocol. All changes to the statistical analyses planned in the protocol and added secondary analyses are documented in Section 3 of this SAP.

Novo Nordisk will be responsible for the statistical analyses and reporting. Data from all sites will be analysed and reported together.

## 2 Statistical considerations

### 2.1 General considerations

No interim analyses or other analyses of unmasked or between group data will be performed before the database is locked. If necessary, a statistical analysis plan (SAP) may be written in addition to the protocol, including a more technical and detailed elaboration of the statistical analyses. The SAP will be finalised before database lock.

Data from all trial sites will be analysed and reported together. If an assessment has been made both at screening and randomisation, and if not otherwise specified, the value from the randomisation visit will be used as the baseline value. If an assessment is missing at randomisation, but available at screening, then the screening value will be used as the baseline value.

Laboratory values below the lower limit of quantification (LLOQ) will be set to  $\frac{1}{2}$ LLOQ.

Results from a statistical analysis will be presented by the estimated treatment contrasts at week 40 with associated two-sided 95% confidence intervals and p-values corresponding to two-sided tests of no difference if not otherwise specified.

The two principal comparisons presented from a statistical analysis are

- s.c. semaglutide 0.5 mg versus dulaglutide 0.75 mg
- s.c. semaglutide 1.0 mg versus dulaglutide 1.5 mg

### 2.2 Primary estimand

The primary objective of the trial is to compare the effect of once-weekly dosing of two dose levels of s.c. semaglutide (0.5 mg and 1.0 mg) versus once-weekly dosing of each of the two dose levels of s.c. dulaglutide (0.75 mg and 1.5 mg) on glycaemic control in subjects with T2D on a background treatment with metformin.

The primary estimand will be:

- de-jure treatment difference at week 40 for all randomised subjects if all subjects adhered to treatment and did not initiate antidiabetic rescue medication

This estimand assesses the glycaemic benefit a future subject is expected to achieve if he/she initiates and continues treatment with s.c. semaglutide as compared to dulaglutide. It is considered a clinically relevant estimand as it provides information to treating clinicians about the expected glycaemic efficacy of s.c. semaglutide for purposes of treating individual subjects with T2D.



## 2.3 Missing data considerations

Since both semaglutide and dulaglutide are GLP-1 RA, it is reasonable to assume that missing data in both arms will be similar in timing, extent and reason. Based on the phase 2 semaglutide dose finding trial (NN9535-1821) and the slower dose escalation implemented in this trial, the rate of discontinuing treatment prematurely or initiating rescue medication on top of trial product is expected to be maximum 25% and similar across treatment arms after 40 weeks of treatment.

Since efficacy for both semaglutide and dulaglutide have been shown, missing data due to ineffective therapy is not anticipated to be a notable issue. However, some missing data due to AEs is expected in both treatment arms because of gastrointestinal related AEs leading to premature treatment discontinuation primarily during initiation and dose escalation.

Descriptive summaries and graphical representation of extent, reason(s) for and pattern of missing data will be presented by treatment arm.

## 2.4 Sample size calculation

The primary endpoint is change from baseline in HbA<sub>1c</sub> after 40 weeks of treatment. For HbA<sub>1c</sub>, both non-inferiority and subsequently superiority are planned to be tested at each dose level (semaglutide 0.5 mg versus dulaglutide 0.75 mg and semaglutide 1.0 mg versus dulaglutide 1.5 mg). The confirmatory secondary endpoint is change from baseline in body weight after 40 weeks of treatment. For body weight, superiority is planned to be tested at each dose level.

The sample size calculation is based on jointly meeting four out of the six pre-specified confirmatory hypotheses shown in [Figure 1](#). The closed testing procedure described in Bretz et al<sup>[1]</sup> is used to control the overall type-1 error at a two-sided 5% level. The four hypotheses are:

- HbA<sub>1c</sub> non-inferiority of semaglutide 0.5 mg versus dulaglutide 0.75 mg (margin of 0.4%)
- Body weight superiority of semaglutide 0.5 mg versus dulaglutide 0.75 mg
- HbA<sub>1c</sub> non-inferiority of semaglutide 1.0 mg versus dulaglutide 1.5 mg (margin of 0.4%)
- Body weight superiority of semaglutide 1.0 mg versus dulaglutide 1.5 mg

The sample size is calculated using the calcPower function in the R package, gMCP<sup>[2]</sup> using 10000 simulations. All of the six pre-specified confirmatory hypotheses are assumed to be independent. Since positive correlations are expected, the assumption of independence is viewed as conservative.

Furthermore, the sample size calculation is based on the following assumptions:

Based on the phase 2 s.c. semaglutide dose-finding trial (NN9535-1821), the standard deviation for HbA<sub>1c</sub> is assumed to be 1.1% and the standard deviation for body weight is assumed to be 4 kg.

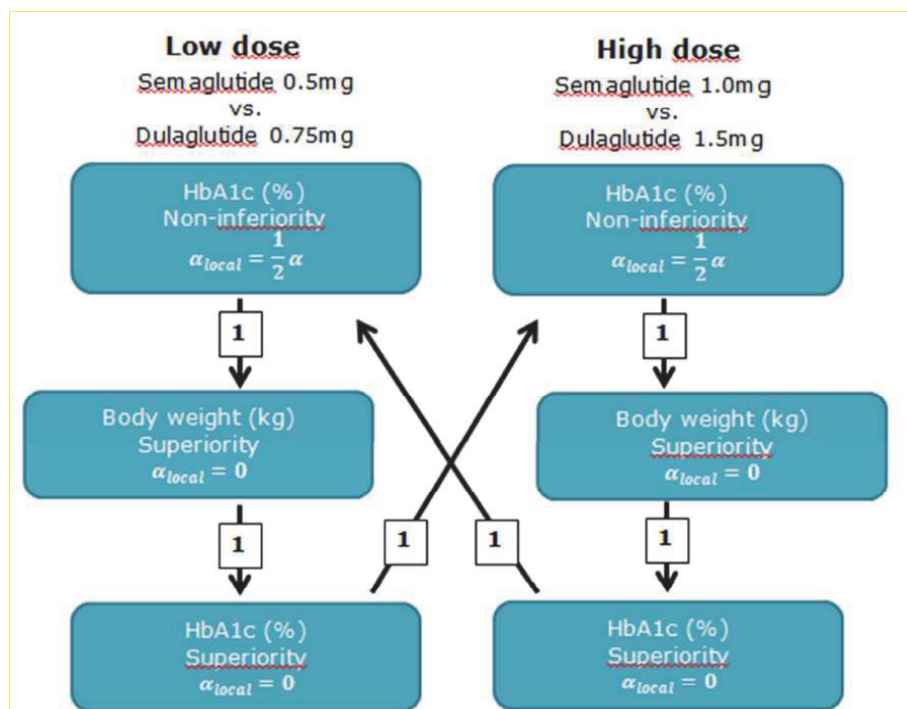
The assumed treatment difference in HbA<sub>1c</sub> of semaglutide relative to dulaglutide at week 40 within both dose levels is zero. The assumed treatment difference in body weight of semaglutide relative to dulaglutide at week 40 within both dose levels is 1.5 kg. A 50% smaller effect on body weight is assumed in the 25% of subjects expected to discontinue treatment prematurely or initiate rescue medication on top of trial product. This leads to an adjusted treatment effect of 1.35 kg for body weight, which is the value used in the sample size calculation. Based on oral semaglutide phase 2 results (NN9924-3790), the 50% efficacy retention is viewed as conservative in light of the primary analysis but less so for the in-trial sensitivity analyses that uses all data collected during the trial.

With the above assumptions, allocating 299 subjects to each of the semaglutide and dulaglutide arms yields 90% power to confirm HbA<sub>1c</sub> non-inferiority and body weight superiority between semaglutide and dulaglutide at both dose levels.

Calculated powers for selected individual hypotheses are presented in [Table 1](#). In total 4×299 = 1196 subjects are planned to be randomised.

**Table 1      Calculated powers for individual hypotheses**

Statistical test	HbA <sub>1c</sub> non-inferiority (margin=0.4%)		Body weight superiority		HbA <sub>1c</sub> superiority	
Treatment dose level	High	Low	High	Low	High	Low
Power (%)	99	99	95	95	1.2	1.2



**Figure 1** Graphical illustration of the closed testing procedure

The Type-I error for the six pre-specified confirmatory hypotheses will be controlled in the strong sense using the closed testing procedure in Figure 1. The initial allocation of the overall significance level of  $\alpha=0.05$  (two-sided) is split equally between non-inferiority at the two dose levels. The local significance level ( $\alpha_{local}$ ) will be reallocated if a hypothesis is confirmed according to the weight given by directed edge between nodes (the hypotheses).

## 2.5 Definition of analysis sets

The following analysis sets will be defined:

**Full analysis set (FAS):** includes all randomised subjects exposed to at least one dose of trial product. Subjects in the FAS will contribute to the evaluation “as randomised”.

**Safety analysis set (SAS):** includes all randomised subjects exposed to at least one dose of trial product. Subjects in the SAS will contribute to the evaluation “as treated”.

**Per protocol (PP) analysis set:** includes all subjects in the FAS who fulfil the following criteria:

- Have not violated any inclusion criteria
- Have not fulfilled any exclusion criteria
- Have a non-missing HbA<sub>1c</sub> measurement at screening and/or randomisation

- Is on trial product at week 28 and have at least one non-missing HbA<sub>1c</sub> measurement at or after week 28

Subjects in the PP analysis set will contribute to the analysis “as treated”.

Before data are locked for statistical analysis, a review of all data will take place. Any decision to exclude either a subject or single observations from the statistical analysis is the joint responsibility of the members of the Novo Nordisk study group.

Exclusion of data from analyses will be used restrictively and normally no data should be excluded from the FAS. The subjects or observations to be excluded, and the reasons for their exclusion will be documented and signed by those responsible before database lock. The subjects and observations excluded from analysis sets, and the reason for this, will be described in the clinical trial report.

## **2.6 Data selections and observations periods**

Unless subjects withdraw their informed consent, data collection will continue for the full duration of the trial, which is until the follow-up phone contact (P10).

The data to be used in all analyses will be selected in two steps.

Step 1: The subjects and treatment principle (as treated or as randomised) to be used in the analysis will be selected based on the specified *analysis set*.

Step 2: Data points for subjects used in the analysis set will be selected according to whether or not the data points belongs to the specified *observation period* (In-trial, on-treatment or on-treatment without rescue as defined below). Information collected with onset date outside the observation period will be treated as missing and therefore excluded from the corresponding analysis. For adjudicated events, onset date will be the EAC adjudicated onset date.

- *In-trial*: The in-trial observation period includes observations recorded at or after randomisation (as registered in IWRS) and not after the last subject-investigator contact, which is scheduled to take place 5 weeks after planned last dose of trial product at a follow-up phone contact (P10). For subjects who withdraw their informed consent, the in-trial observation period ends at their date of withdrawal. In the case a subject dies during the trial, the date of death will be the end-date of the in-trial observation period. If a subject is lost to follow-up, the end of his/her in-trial period is defined as the date of the last subject-investigator contact (site or phone visit). Analysis based on this observation period includes data regardless of treatment exposure and/or usage of non-investigational antidiabetic medications. Since non-investigational antidiabetic medications can mask or exaggerate both the efficacy and safety effects, this observation period will be in line with the primary estimand and is considered supportive for both efficacy and safety evaluations.

- *On-treatment*: This observation period represents the time period where subjects are considered treated with trial product. The observation period is a subset of the in-trial observation period and two slightly different data handling rules will be needed to cover all assessments appropriately. For adjudicated events, ECGs and AEs including hypoglycaemic episodes, this observation period will represent information collected while subjects are considered exposed to trial product. This corresponds to information collected up until the follow-up phone contact, which is scheduled to take place 5 weeks after premature treatment discontinuation to reflect five half-lives of s.c. semaglutide including a visit window of +7 days. For the remaining safety and efficacy assessments that are collected up until and not after the end of treatment visit, the follow-up period will not be included in the on-treatment observation period. In line with the primary estimand, the on-treatment observation period will be the primary observation period used in safety evaluations.
- *On-treatment without rescue*: This observation period is a subset of the on-treatment observation period, where subjects do not receive any non-investigational antidiabetic medication (rescue medication). Specifically it includes observations recorded at or after date of first dose of trial product and not after the first occurrence of the following:
  - The last dose of trial product plus the dosing interval
  - Initiation of rescue medication

For subjects who have no post-baseline scheduled assessments available in the on-treatment without rescue period, the baseline value will be carried forward to the first scheduled visit for the associated endpoint to ensure that all randomised subjects will contribute to the statistical analysis. In line with the primary estimand, the on-treatment without rescue observation period will be the primary observation period used in efficacy evaluations.

## 2.7 Primary endpoint

The primary analysis used to estimate the primary estimand will be based on FAS using data from the on-treatment without rescue observation period in a Mixed Model for Repeated Measures (MMRM). A restricted maximum likelihood (REML) will be used. The model will include all post baseline HbA<sub>1c</sub> measurements collected at scheduled visits up to and including week 40 data as dependent variables. The independent effects included in the model will be treatment and country as fixed effects and baseline response as covariate, all nested within visit. An unstructured covariance matrix will be employed for measurements within the same subject, assuming that measurements across subjects are independent. From this model, the two by dose level estimated treatment differences between s.c. semaglutide versus dulaglutide at week 40 will be presented together with associated two-sided 95% confidence intervals and unadjusted two sided p-values for testing non-inferiority and superiority.

The MMRM is a well-established method that accounts for the uncertainty pertaining to missing data. This analysis assumes that the missing data mechanism is missing at random (MAR). Thus,

for a subject who has missing data, MAR assumes a value for the endpoint based on observed data of subjects whose baseline explanatory variables and response up to withdrawal are similar to that of the discontinued subject. Since there is no historical evidence suggesting that subjects discontinuing semaglutide prematurely have better outcome on average than those who remain on treatment, the primary analysis is not expected to bias the estimated HbA<sub>1c</sub> treatment effect for the primary estimand in favour of semaglutide to any important degree. This is based on the s.c. semaglutide phase 2 (NN9535-1821) results and further supported by data from liraglutide clinical trials.

For an overview of statistical analyses to be performed for the primary endpoint HbA<sub>1c</sub>, see [Table 2](#).

**Table 2 Summary of statistical analyses of primary endpoint (HbA<sub>1c</sub>)**

Population	Period	Statistical model	Imputation <sup>1)</sup>	Hypothesis to be tested
<b>Primary analyses</b>				
FAS	On-treatment without rescue medication	MMRM <sup>2)</sup>	Baseline values will be carried forward to week 4 for subjects with all post-baseline values missing.	Non-inferiority
FAS	On-treatment without rescue medication	MMRM <sup>2)</sup>	Baseline values will be carried forward to week 4 for subjects with all post-baseline values missing.	Superiority
<b>Sensitivity analyses</b>				
PP	On-treatment without rescue medication	MMRM	Baseline values will be carried forward to week 4 for subjects with all post-baseline values missing.	Non-inferiority
Complete cases	On-treatment without rescue medication	ANCOVA <sup>4)</sup>	Not applicable	Non-inferiority

FAS	On-treatment without rescue medication	ANCOVA <sup>4)</sup>	Last observation carried forward for subjects with missing value at week 40	Superiority
FAS	On-treatment without rescue medication	Tipping point analysis using an ANCOVA model	Multiple imputation	Non-inferiority
FAS	On-treatment without rescue medication	Tipping point analysis using an ANCOVA model	Multiple imputation	Superiority
FAS	In-trial	Retrieved dropout analysis, using an ANCOVA model <sup>3)</sup>	Multiple imputation	Superiority

<sup>1)</sup> If a subject is missing baseline assessment, then the screening assessment will be used as the baseline assessment. If no measurement(s) have been obtained, at or prior to randomisation, the baseline value will be left missing.

<sup>2)</sup> MMRM model will be used to estimate the efficacy (de-jure) estimand

<sup>3)</sup> Retrieved dropout analysis, using an ANCOVA model, will be used to estimate the effectiveness (de-facto) estimand

<sup>4)</sup> Baseline and week 40 measurements will be used in the analysis

## 2.8 Hypotheses tested for the primary endpoint

For HbA<sub>1c</sub>, the following two confirmatory hypotheses are planned to be tested at each dose level comparing; (i) semaglutide 0.5 mg versus dulaglutide 0.75 mg and (ii) semaglutide 1.0 mg versus dulaglutide 1.5 mg with mean treatment difference defined as  $\mu = (\text{semaglutide} - \text{dulaglutide})$ :

- Non-inferiority using a non-inferiority margin of 0.4  
–  $H_0: \mu \geq 0.4\%$  against  $H_a: \mu < 0.4\%$
- Superiority  
–  $H_0: \mu \geq 0.0\%$  against  $H_a: \mu < 0.0\%$

Operationally, the hypotheses will be assessed using two-sided p-values.

## 2.9 Multiplicity and criteria for confirming hypotheses

The Type-I error for testing the six confirmatory hypotheses related to HbA<sub>1c</sub> and body weight endpoints will be preserved in the strong sense at 5% (two-sided) using the weighted Bonferroni-based closed testing procedure described in Bretz et al<sup>[1]</sup> and outlined in Figure 1. First the 2 non-inferiority hypotheses at each dose level will be tested each at its initial allocated local significance

level of 0.025%. If a non-inferiority hypothesis is confirmed, the local significance level will be reallocated according to the edge going out of the confirmed hypothesis as specified in [Figure 1](#). Each of the following hypotheses will be tested at their local significance level ( $\alpha$ - local). This process will be repeated until no further hypothesis can be confirmed.

Non-inferiority and/or superiority will be considered confirmed if the mean treatment difference is supporting the corresponding hypothesis and the two-sided p-value is strictly below its local two-sided significance level as defined by the closed testing procedure in [Figure 1](#). This is equivalent to using a one-sided p-value and a one-sided 2.5% overall significance level in the closed testing procedure.

## **2.10 Sensitivity analyses**

The aim of the below pre-specified sensitivity analyses is to explore the impact of departures from the missing data assumption made in the primary analysis of HbA<sub>1c</sub> and the confirmatory secondary analysis of body weight (see section [2.11.1](#)). This is consistent with European Medicines Agency (EMA) recommendations <sup>[3]</sup> and with a report from the US National Research Council <sup>[4]</sup>. Since conservatism (i.e. avoiding bias in favour of semaglutide) depends on the context, the sensitivity analyses are targeted to whether non-inferiority or superiority is being tested.

### **2.10.1 Tipping point analyses**

The tipping point analyses explore the validity of the conclusions in the trial, where missing data will be handled using multiple imputation assuming that missing data is missing at random (MAR).

Tipping-point analysis (pattern mixture model based) is based on the FAS using the 'on-treatment without rescue medication' observation period. In this analysis, subjects from the semaglutide group with missing observations will be given a penalty, i.e., it is assumed that subjects with missing observations who are randomised to semaglutide will receive a treatment that is less beneficial than subjects with observed values who are randomised to semaglutide. The idea is to gradually increase the penalty to evaluate at which level the conclusion of the analyses in terms of statistical significance is changed. It can be used to evaluate the robustness of statistical significance when the extent of missing data is reasonable <sup>[5]</sup>.

Intermittent missing values, within the given observation period, are imputed using a Markov Chain Monte Carlo (MCMC) method, in order to obtain a monotone missing data pattern. This imputation is done for each of the four treatment groups separately, defined by randomised treatment, and 500 datasets will be generated. These 500 datasets have a monotone missing data pattern and will be used as data foundation for imputation of remaining monotone missing values.

For each of the 500 datasets (MCMC imputed), a sequential conditional regression approach for imputing monotone missing values at planned visits will be implemented starting with the first visit



after baseline and sequentially continuing to the last planned visit at week 40. An ANCOVA model used to impute missing values at each planned visit will be fitted for each of the four treatment groups using observed data. The model will include baseline and post-baseline values observed prior to the visit in question as covariates. The estimated parameters, and their variances, from this model are used to impute missing values for subjects in each treatment groups.

For each of the complete 500 datasets, penalty values are added stepwise to the imputed change from baseline at week 40, followed by performing an ANCOVA. The addition of the penalty values and subsequent analysis steps should be repeated with increasing penalty values until a significant result in the corresponding superiority and non-inferiority analyses are no longer significant. The tipping point will occur at the penalty level, at which the magnitude of efficacy reduction in subjects with missing data creates a shift in the treatment effect of semaglutide from being statistically significantly better than dulaglutide to being non-statistically significantly better for the superiority test and similarly for the non-inferiority test. There will be one tipping point per dose level (Sema 0.5 mg vs. Dula 0.75 mg and Sema 1.0 mg vs. Dula 1.5 mg) and hypotheses test (superiority/ non-inferiority), e.g. four estimates in total.

For more technical details, see SPS<sup>[6]</sup>, where also the seeds used when data is generated, can be found.

### **2.10.2 Sensitivity analyses addressing non-inferiority**

In support of non-inferiority testing, the below two sensitivity analyses will be performed. These sensitivity analyses only include a subset of all randomised subjects so the integrity of randomisation may not be maintained. Therefore, while the below two analyses generally are conservative for testing non-inferiority, the inherent risk for bias in any direction cannot be excluded.

- *PP analysis* – the statistical analysis will be the same as the primary MMRM based analysis but it will be based on the PP analysis set and the on-treatment without rescue observation period.
- *Complete case analysis* – includes subjects in the FAS who do not have their endpoint imputed in the primary analysis. The change from baseline in HbA<sub>1c</sub> at week 40 will be analysed by a linear normal model (analysis of covariance (ANCOVA)) with treatment and country as fixed effects and baseline HbA<sub>1c</sub> as a covariate.

### **2.10.3 Other sensitivity analyses**

The following secondary estimand will be defined:

- de-facto treatment difference at week 40 for all randomised subjects

This estimand assesses the average effect in a future population that results from treatment with s.c. semaglutide plus antidiabetic rescue medication(s) as compared to treatment with dulaglutide plus antidiabetic rescue medication(s). Interpretation of this estimand depends on whether the use of antidiabetic rescue medication and treatment adherence in this trial reflects clinical practice. The defacto estimand will be estimated from the below analysis:

- *Retrieved dropout analysis* – This in-trial treatment policy analysis will be based on the FAS using the in-trial observation period. Missing data will be imputed within the same group defined, not only by the randomised treatment (semaglutide/ dulaglutide), but also by the status of treatment completion (still on randomised treatment at week 40 yes/no) (8 groups in total). It is hereby assumed that the likely values of what the missing data would have been if available are best described by information from subjects who at week 40 are similar in terms of randomised treatment and treatment completion status. For each group missing values will be imputed using a MCMC method, as described in section 2.10.1. For each of the 500 datasets, an ANCOVA is applied to data from baseline to end-of-treatment, week 40. The model use treatment and country as factors and baseline value as covariate. The estimates and standard deviations for the 500 data sets are pooled to one estimate and associated standard deviation using Rubin's rule[7]. From these pooled estimates the confidence interval for the treatment differences and the associated p-value are calculated.

The last sensitivity analysis will be:

- *Last observation carried forward (LOCF) analysis* – This analysis will be based on the FAS using the on-treatment without rescue observation period with missing data imputed by LOCF. Based on the complete data set, the change from baseline in HbA<sub>1c</sub> at week 40 will be analysed by a linear normal model (ANCOVA) with treatment and country as fixed effects and baseline HbA<sub>1c</sub> as a covariate.

#### **2.10.4 Assessment of sensitivity analyses**

The results from the sensitivity analysis will be collectively used to interpret the confirmatory trial conclusions on HbA<sub>1c</sub> and body weight, and in particular evaluate the impact of the MAR assumptions. No absolute criteria will be defined as to when a sensitivity analysis can be defined to have confirmed the robustness of the conclusions. Due to the large number of the sensitivity analyses and their inherent conservative nature, it is not considered a requirement that all confirmatory hypotheses are confirmed across all the sensitivity analyses. The results of the sensitivity analysis will be discussed in the clinical trial report with the aim to use the sensitivity results in totality to evaluate the credibility of the confirmatory trial conclusions.

## 2.11 Secondary endpoints

### 2.11.1 Confirmatory secondary endpoint

The confirmatory secondary endpoint is change from baseline to week 40 in body weight (kg). This endpoint will be analysed in the same type of model as the primary endpoint, except with baseline body weight as a covariate instead of baseline HbA<sub>1c</sub>. From this model the two by dose level estimated treatment differences between semaglutide versus dulaglutide will be presented at week 40 together with associated two-sided 95% confidence intervals and unadjusted two sided p-values. The same sensitivity analyses as pre-specified for testing superiority for the primary HbA<sub>1c</sub> endpoint will also be performed to evaluate the robustness of the body weight superiority conclusions.

For an overview of statistical analyses to be performed for the confirmatory secondary endpoint, body weight, see [Table 3](#).

**Table 3 Summary of sensitivity analyses of body weight**

Population	Period	Imputation	Statistical model
FAS	On-treatment without rescue medication	Last observation carried forward for subjects with missing value at week 40	ANCOVA <sup>1)</sup>
FAS	On-treatment without rescue medication	Multiple imputation	Tipping point analysis
FAS	In-trial	Multiple imputation	Retrieved dropout analysis, using an ANCOVA model

<sup>1)</sup> Baseline and week 40 measurements will be used in the analysis

### 2.11.2 Confirmatory secondary hypothesis

For body weight, the following confirmatory hypothesis will be tested at each dose level comparing; (i) semaglutide 0.5 mg versus dulaglutide 0.75 mg and (ii) semaglutide 1.0 mg versus dulaglutide 1.5 mg with mean treatment difference defined as  $\mu = (\text{semaglutide} - \text{dulaglutide})$ :

- Superiority
  - $H_0: \mu \geq 0.0\%$  against  $H_a: \mu < 0.0\%$

Superiority will be considered confirmed if the corresponding two-sided p-value is strictly below its local two-sided significance level as defined by the closed testing procedure in [Figure 1](#).

## 2.12 Supportive secondary endpoints

### 2.12.1 Efficacy endpoints

The supportive secondary efficacy endpoints will be presented based on FAS using the on-treatment without rescue observation period as the key observation period with the in-trial observation period being supportive.

Endpoints include change from baseline to week 40 in:

- Relative change in body weight (%)
- FPG\*
- SMPG, 7-point profile:
  - Mean 7-point profile, defined as the area under the profile, calculated using the trapezoidal method, divided by the measurement time
  - Mean post prandial increment (over all meals)
- Fasting blood lipids (total cholesterol, LDL cholesterol, HDL cholesterol, triglycerides)
- BMI and waist circumference
- Systolic and diastolic blood pressure\*
- Patient reported outcomes
  - SF-36v2™
  - DTSQ\*

The above continuous endpoints will be analysed separately using a similar model as for the primary endpoint but with the associated baseline value as a covariate. Fasting blood lipids profile endpoints will be log-transformed prior to analysis including also the relevant log-transformed baseline value used as a covariate.

#### Subjects who after 40 weeks treatment achieve (yes/no)

- HbA<sub>1c</sub> <7.0% (53 mmol/mol) ADA target
- HbA<sub>1c</sub> ≤6.5% (48 mmol/mol) AACE target\*
- Weight loss ≥5%
- Weight loss ≥10%
- HbA<sub>1c</sub> <7.0% (53 mmol/mol) without severe or BG confirmed symptomatic hypoglycaemia episodes and no weight gain
- HbA<sub>1c</sub> reduction ≥1%
- Weight loss ≥3%
- HbA<sub>1c</sub> reduction ≥1% and weight loss ≥3%

The above eight binary endpoints will be analysed using a logistic regression model with treatment and region as fixed effects and baseline response as covariate (i.e. baseline HbA<sub>1c</sub> for binary HbA<sub>1c</sub>

endpoints, baseline weight for weight endpoints and both baseline HbA<sub>1c</sub> and baseline weight for the binary endpoint that combines both parameters). To account for missing data, the analysis will be made using a sequential multiple imputation approach as described below:

- Multiple imputed data sets (500) will be created in which missing values for the underlying continuous assessments are imputed by treatment group assuming MAR and as described in section 2.7.
- The binary endpoint will be created for each of the 500 complete data sets
- Each of the created complete data set will be analysed with the logistic model and inference will be drawn using Rubin's rule <sup>[7]</sup>.

## **PRO outcomes**

The PRO outcomes endpoints that will be analysed at week 40 are:

- PRO questionnaire outcome DTSQs (individual items and treatment satisfaction score (6 of the 8 items summed)), and
- PRO questionnaire outcome SF-36v2™

The above continuous endpoints will be analysed separately using a similar model as for the primary endpoint but with the associated baseline value as a covariate.

## **7-point profile (SMPG)**

Subjects will be asked to perform SMPG measurements before and 90 minutes after breakfast, lunch, dinner, respectively, and at bedtime.

The endpoints from the 7-point profiles that will be analysed at week 40 are:

- Mean of the 7-point profile, defined as the area under the profile, calculated using the trapezoidal method, divided by the measurement time
- Mean increment over all meals

The mean of the 7-point profile and the mean of the post prandial increments at week 40 will be analysed separately using a similar model as for the primary endpoint but with the associated baseline value as a covariate.

### **2.12.2 Safety endpoints**

All safety endpoints will be evaluated based on SAS and the on-treatment observation period as the primary observation period with the in-trial observation period being supportive if not otherwise specified.

The following endpoints are used to support the safety objectives:

- Number of TEAEs
- Number of treatment emergent severe or BG confirmed symptomatic hypoglycaemia episodes
- Treatment emergent severe or BG confirmed symptomatic hypoglycaemia (yes/no)

Change from baseline to week 40 in:

- Lipase
- Amylase
- Pulse

The above continuous endpoints will be analysed separately using a similar model as for the primary endpoint but with the associated baseline value as covariate. Lipase and amylase endpoints will be log-transformed prior to analysis including also the relevant log-transformed baseline value as covariate.

The following laboratory assessments will be summarised descriptively:

- Haematology
- Biochemistry
- Calcitonin

The following categorical safety evaluations will be summarised descriptively:

- ECG category
- Physical examination

## **Calcitonin**

In addition to the continuous summaries, calcitonin will be displayed in terms of the number of subjects (N), the percentage of subjects (%) and the event rate per 100 years of exposure (R). The below criteria are defined for categorical tabulations. Summaries will be presented for all subjects and by gender.

- Persistent (all post baseline measurements)
- From < upper normal limit (UNL) to persistently  $\geq$  UNL
- From <UNL to persistently  $\geq 1.5$  UNL
- From <UNL to persistently  $\geq 20$  ng/L
- From <UNL to persistently  $\geq 50$  ng/L
- From <20 ng/L to persistently  $\geq 20$  ng/L
- From <50 ng/L to persistently  $\geq 50$  ng/L

### 2.12.3 Adverse Events

All AEs will be coded using the most recent version of the Medical Dictionary for Regulatory Activities (MedDRA) coding.

A TEAE is defined as an AE with onset in the on-treatment period (see definition of observation period in section 2.6).

TEAEs will be summarised in terms of the number of subjects with at least one event (N), the percentage of subjects with at least one event (%), the number of events (E) and the event rate per 100 patient years of observation time (R) for the on-treatment observation period. Supportive summaries of AEs will be made for the in-trial observation period.

#### **Classification of Hypoglycaemia:**

Hypoglycaemic episodes will be summarised for the SAS and the on-treatment observation period. No supportive summaries will be made based on the in-trial observation period and episodes with onset date after the on-treatment observation period will be reported in listings only.

Treatment emergent: hypoglycaemic episodes will be defined as treatment emergent if the onset is in the on-treatment period (see definition of observation period in section 2.6)

Nocturnal hypoglycaemic episodes: are episodes occurring between 00:01 and 05.59 both inclusive.

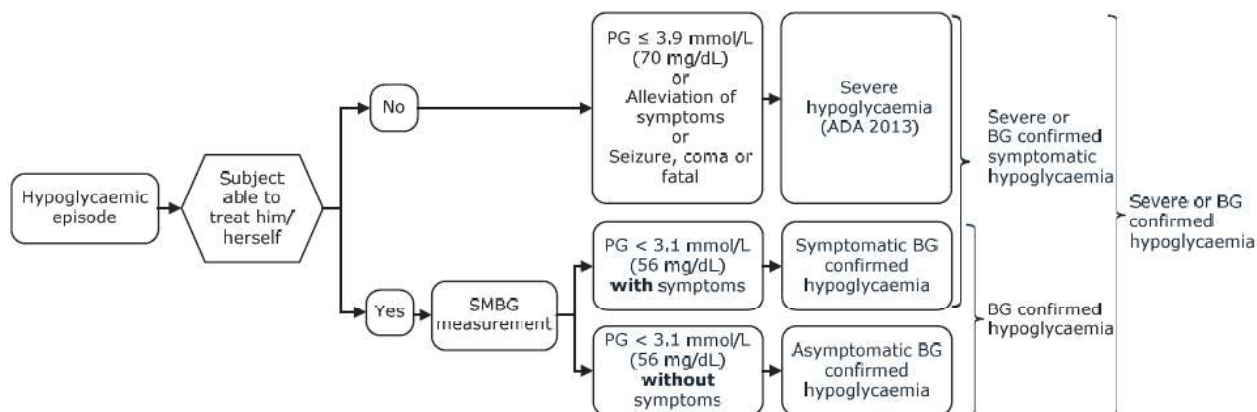
Hypoglycaemic episodes are classified according to the Novo Nordisk classification of hypoglycaemia (see [Figure 2](#)) and the ADA classification of hypoglycaemia (see [Figure 3](#)).

#### **Novo Nordisk classification of hypoglycaemia**

In normal physiology, symptoms of hypoglycaemia occur below a plasma glucose level of 3.1 mmol/L (56 mg/dL) <sup>[8]</sup>. Therefore, Novo Nordisk has included hypoglycaemia with plasma glucose levels below this cut-off point in the definition of BG confirmed hypoglycaemia.

Novo Nordisk uses the following classification (see [Figure 2](#)) in addition to the ADA classification:

- Severe or BG confirmed symptomatic hypoglycaemia: An episode that is severe according to the ADA classification<sup>[9]</sup> or BG confirmed by a plasma glucose value <3.1 mmol/L (56 mg/dL) with symptoms consistent with hypoglycaemia.



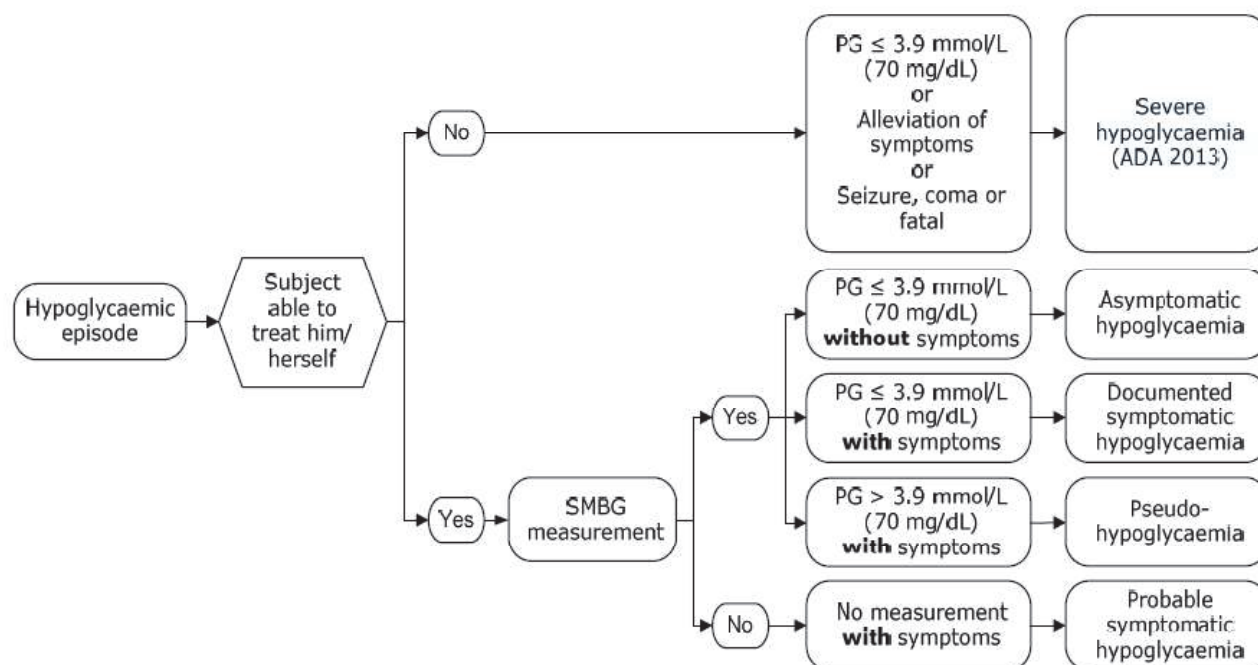
Note: Glucose measurements are performed with capillary blood calibrated to plasma equivalent glucose values

**Figure 2 Novo Nordisk classification of hypoglycaemia**

#### ADA classification<sup>[9]</sup> of hypoglycaemia

- Severe hypoglycaemia: An episode requiring assistance of another person to actively administer carbohydrate, glucagon, or take other corrective actions. Plasma glucose concentrations may not be available during an event, but neurological recovery following the return of plasma glucose to normal is considered sufficient evidence that the event was induced by a low plasma glucose concentration.
- Asymptomatic hypoglycaemia: An episode not accompanied by typical symptoms of hypoglycaemia, but with a measured plasma glucose concentration  $\leq 3.9$  mmol/L (70 mg/dL).
- Documented symptomatic hypoglycaemia: An episode during which typical symptoms of hypoglycaemia are accompanied by a measured plasma glucose concentration  $\leq 3.9$  mmol/L (70 mg/dL).
- Pseudo-hypoglycaemia: An episode during which the person with diabetes reports any of the typical symptoms of hypoglycaemia with a measured plasma glucose concentration  $> 3.9$  mmol/L (70 mg/dL) but approaching that level.
- Probable symptomatic hypoglycaemia: An episode during which symptoms of hypoglycaemia are not accompanied by a plasma glucose determination but that was presumably caused by a plasma glucose concentration  $\leq 3.9$  mmol/L (70 mg/dL).





Note: Glucose measurements are performed with capillary blood calibrated to plasma equivalent glucose values

### Figure 3 ADA classification of hypoglycaemia

Data on treatment emergent hypoglycaemic episodes are presented in terms of the number of subjects with at least one episode, the percentage of subjects with at least one episode (%), the total number of episodes and the episode rate per 100 years of exposure. Summaries of treatment emergent hypoglycaemic episodes will be presented as an overview including all episodes and episodes by severity.

### Number of severe or BG confirmed symptomatic hypoglycaemic episodes

The number of treatment emergent severe or BG confirmed symptomatic hypoglycaemic episodes will be analysed using a negative binomial regression model with a log-link function and the logarithm of the time period covered by the subject's on-treatment observation period as offset. The model will include factors for treatment and region as fixed factors and baseline HbA<sub>1c</sub> as covariate.

### Severe or BG confirmed symptomatic hypoglycaemic episodes (yes/no)

The binary endpoint indicating whether a subject has at least one treatment emergent severe or BG confirmed symptomatic hypoglycaemic episodes will be analysed using a logistic regression model with treatment and region as fixed factors and baseline HbA<sub>1c</sub> as covariate.

## **2.13 Health economics and/or patient reported outcomes**

The PRO questionnaires, SF-36v2™ and DTSQs, derived endpoints for overall scores and domains will be analysed separately using a similar model as for the primary endpoint but with the associated baseline value as covariate.

### 3 Changes to the statistical analyses planned in the protocol

The main analyses were described in the protocol for trial NN9535-4216. More detailed descriptions of endpoints and analyses are provided in this SAP. Only key changes and major additions to the protocol are described below.

Changes from protocol NN9535-4216:

- ‘and considered supportive of efficacy evaluations’ has been deleted in section 2.6 “Data selections and observations periods”. No efficacy evaluation will be done in the ‘On-treatment’ observation period.
- To get an overview of statistical analyses to be performed for the primary (HbA1c) and confirmatory secondary (body weight) endpoints, table 2 and table 3 have been added in respective section, 2.7 and 2.11.
- Pattern-mixture model using multiple imputation addressing both non-inferiority and superiority has been deleted. A pattern-mixture model using multiple imputation a ‘Tipping point analysis’ will be performed instead of the comparator-based analysis described in protocol. See section 2.10.1.
- To evaluate the robustness of the primary analyses, a ‘Retrieved dropout analysis’, described in section 2.10.3, will be performed instead of the ‘In-trial analysis’, which was stated in the protocol.
- Relative change from baseline in body weight has been added in section 2.12.1.
- Number of multiple imputed data sets has been changed from 100 to 500, see section 2.12.1.
- “(nominal alpha=0.05)” in section 2.4, 2.7, 2.9 and 2.11.2 has been deleted to avoid confusion. The overall significance level is  $\alpha=0.05$  (two-sided).

## 4 References

1. Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Stat Med* 2009; 28(4):586-604.
2. Rohmeyer, K. and Klinglmueller, F. gMCP: Graph Based Multiple Test Procedures. R package version 0.8-8. 3 Oct 2014.
3. European Medicines Agency Committee for Medicinal Products for Human Use (CHMP). Guideline on Missing Data in Confirmatory Clinical Trials (EMA/CPMP/EWP/1776/99 Rev. 1). EMA . 2 Jul 2010.
4. National Academy of Sciences (NAS). The Prevention and Treatment of Missing Data in Clinical Trials. Washington D.C.: The National Academies Press, 2010.
5. Xu Yan , Shiowjen Lee & Ning Li (2009) Missing Data Handling Methods in Medical Device Clinical Trials, *Journal of Biopharmaceutical Statistics*, 19:6, 1085-1098, <http://dx.doi.org/10.1080/10543400903243009>
6. 4216 Statistical Programming Specification, novoDOCs ID number: 003660253
7. Little RJA, Rubin DB. Statistical analysis with missing data. New York: John Wiley & Sons, 1987.
8. Schwartz NS, Clutter WE, Shah SD, Cryer PE. Glycemic thresholds for activation of glucose counter regulatory systems are higher than the threshold for symptoms. *J Clin Invest* 1987; 79(3):777-781.
9. Seaquist ER, Anderson J, Childs B, Cryer P, Dagogo-Jack S, Fish L et al. Hypoglycemia and diabetes: a report of a workgroup of the American Diabetes Association and the Endocrine Society. *Diabetes Care* 2013; 36(5):1384-1395.